THE MODIFIED CRAMÉR-VON MISES GOODNESS-OF-FIT CRITERION FOR TIME SERIES

T.W. Anderson
M.A. Stephens

TECHNICAL REPORT No. 478
JANUARY 17, 1994

Prepared Under Contract
N00014-92-J-1264 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065

THE MODIFIED CRAMÉR-VON MISES GOODNESS-OF-FIT CRITERION FOR TIME SERIES

T.W. Anderson
M.A. Stephens

TECHNICAL REPORT No. 478
JANUARY 17, 1994

Prepared Under Contract
N00014-92-J-1264 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

And The
NATIONAL SCIENCE FOUNDATION
Grant No. DMS91-04432

Professor Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065

(12)

The Modified Cramér-von Mises Goodness-of-Fit Criterion for Time Series

T. W. Anderson and M. A. Stephens

1. Introduction

P. C. Mahalanobis may be best known among mathematical statisticians for the "Mahalanobis distance," which in one form is* $(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)$, where μ_1 and μ_2 are the vectors of means in two populations or distributions and Σ is a covariance matrix common to the two. It is considered a measure of the difference between the two distributions. When Σ is known and μ_1 and μ_2 are unknown and estimated by sample means $\bar{x}_1$ and $\bar{x}_2$, respectively, $(\bar{x}_1-\bar{x}_2)'\Sigma^{-1}(\bar{x}_1-\bar{x}_2)$, also known as a Mahalanobis distance, is an estimator of the first form. If $\bar{x}_1$ and $\bar{x}_2$ are normally distributed and hence $\bar{x}_1-\bar{x}_2$ is normally distributed, the Mahalanobis distance is a quadratic form in normal variables. When $\bar{x}_1$ and $\bar{x}_2$ have covariance matrices Σ and $\mu_1=\mu_2$, the Mahalanobis distance is proportional to a χ^2 variable with number of degrees of freedom equal to the number of components of μ_1, μ_2, $\bar{x}_1$, and $\bar{x}_2$. Another Mahalanobis distance is $(\bar{x}_1-\mu_2)'\Sigma^{-1}(\bar{x}_1-\mu_2)$, which is appropriate if an unknown μ_1 is estimated by a sample mean $\bar{x}_1$.

In this paper we consider somewhat analogous quadratic forms in normal variables when the dimensionality is infinite. Then the quadratic forms are distributed as infinite weighted sums of χ^2-variables. These come about as goodness-of-fit criteria for a hypothesis that a cumulative distribution function is a specified one or that two cdf's are the same. Such criteria also arise for goodness-of-fit tests for standardized spectral distributions.

As examples, we give tables of the distribution of the criterion for testing the hypothesis that a stationary stochastic process is a given moving average process of

*The quadratic form should be called "Mahalanobis distance squared", but the third word is usually omitted.

order 1 and for testing the hypothesis that it is a specified autoregressive process of order 1. Two methods are described for calculating the distribution. Either method is appropriate for calculating the distribution of the criterion for testing the hypothesis that a process is a stationary process whose standardized spectral density or distribution is a specified one.

Test for a given probability distribution. Let $x_1, \cdots, x_n$ be n observations from a distribution with cdf $F(x)$ and let the empirical distribution of the sample be $F_n(x) = k/n$ if k of the x_i's are not greater than x. To test the null hypothesis $F(x) = F_0(x)$, where $F_0(x)$ is specified, the Cramér-von Mises statistic is

$$W_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 dF_0(x). \tag{1}$$

When $F_0(x)$ is continuous and $F(x) = F_0(x)$, $\sqrt{n}[F_n(x) - F_0(x)]$, considered as a stochastic process with time parameter x, converges weakly to a Gaussian process with covariance function $\min[F_0(x), F_0(y)] - F_0(x)F_0(y)$, and W_n^2 has a limiting distribution as a quadratic functional of the process. If we make the monotonic tansformation $u = F_0(x)$ and $u_i = F_0(x_i)$, $i = 1, \cdots, n$, then the u_i can be considered as observations from the uniform distribution on $[0, 1]$. Let $H_n(u) = F_n[F^{-1}(u)]$. Then $\sqrt{n}[H_n(u) - u]$ converges weakly to a Gaussian process, say $J(u)$, with covariance function

$$\mathcal{E}J(u)J(v) = h(u, v) = \min(u, v) - uv. \tag{2}$$

In fact, the limiting process $J(u)$ is the Brownian bridge.

Suppose $F_1(x)$ and $F_2(x)$ are two cdf's, and $F_{n_1}(x)$ and $F_{n_2}(x)$ are the empirical cdf's of samples of size n_1 and n_2 obtained from the two populations, respectively. Then

$$\frac{n_1 n_2}{(n_1 + n_2)^2} \int_{-\infty}^{\infty} [F_{n_1}(x) - F_{n_2}(x)]^2 d[n_1 F_{n_1}(x) + n_2 F_{n_2}(x)] \tag{3}$$

is a criterion for testing the hypothesis $F_1(x) = F_2(x)$. When the hypothesis is true, (3) has the same limiting distribution as (1). The statistic (3) might be considered as analogous to the Mahalanobis distance $(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$, where $\boldsymbol{S}$ is an estimator of $\boldsymbol{\Sigma}$.

Test for a given spectral distribution. Cramér-von Mises tests can be given for testing hypotheses about spectral distributions of stationary stochastic processes. In

many situations questions arise about the pattern of dependence; these questions pertain to the autocorrelations or equivalently to the Fourier transform of the autocorrelations, which we term the *standardized spectral density.*

Consider a stationary stochastic process $\{y_t\}$, $t = \cdots, -1, 0, 1, \cdots$ with $\mathcal{E}y_t = 0$, autocovariance function $\mathcal{E}y_t y_{t+h} = \sigma(h)$, $h = \cdots, -1, 0, 1, \cdots$ and autocorrelation function $\rho_h = \sigma(h)/\sigma(0)$, $h = \cdots, -1, 0, 1, \cdots$. We define the standardized spectral density as

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \rho_h \cos \lambda h, \quad -\pi \leq \lambda \leq \pi. \tag{4}$$

Note that the coefficients of the trigonometric functions are the autocorrelations, not the autocovariances. The Fourier transform of the standardized spectral density is

$$\rho_h = \int_{-\pi}^{\pi} f(\lambda) \cos \lambda h, \quad h = \cdots, -1, 0, 1 \cdots. \tag{5}$$

Knowledge of the standardized spectral density is equivalent to knowledge of the autocorrelations. Since $f(\lambda) = f(-\lambda)$, we define the standardized spectral distribution as

$$F(\lambda) = 2 \int_0^{\lambda} f(v) d(v) = \frac{1}{\pi} \left(\lambda + 2 \sum_{h=1}^{\infty} \rho_h \frac{\sin \lambda h}{h} \right). \tag{6}$$

Note that $F(\pi) = 1$; the standardized spectral distribution has the properties (nonnegative increments) of a probability distribution on $[0, \pi]$.

If the sample is $y_1, \cdots, y_T$, the sample autocovariance sequence may be defined as

$$c_h = c_{-h} = \frac{1}{T-h} \sum_{t=1}^{T-h} y_t y_{t+h}, \quad h = 0, 1, \cdots, T-1, \tag{7}$$

Note that c_h here is an unbiased estimator of $\sigma(h)$. The reason for using the unbiased estimator instead of the biased estimator (that is, $\sum_{t=1}^{T-h} y_t y_{t+h}/T$ as in Anderson (1993)) is that then the asymptotic theory serves as a better approximation to the small sample behavior. The sample autocorrelation sequence is defined as $r_h = r_{-h} = c_h/c_0$, $h = 0, 1, \cdots, T-1$; the standardized sample spectral density is

$$I_T(\lambda) = \frac{1}{2\pi c_0} \left| \sum_{t=1}^{T} y_t e^{i\lambda t} \right| = \frac{1}{2\pi} \sum_{h=-(T-1)}^{T-1} r_h \cos \lambda h, \quad -\pi \leq \lambda \leq \pi; \tag{8}$$

and the standardized sample spectral distribution function is

$$F_T(\lambda) = 2\int_0^\lambda I_T(v)dv = \frac{1}{\pi}\left(\lambda + 2\sum_{h=1}^{T-1} r_h \frac{\sin \lambda h}{h}\right). \tag{9}$$

To test the hypothesis that $F(x) = F_0(x)$, where $F_0(x)$ is completely specified, one can use the Cramér-von Mises statistic

$$\frac{T}{2\pi G^2(\pi)}\int_0^\pi [F_T(\lambda) - F_0(\lambda)]^2 f_0^2(\lambda)d\lambda. \tag{10}$$

Consider the discrepancy

$$\sqrt{T}[F_T(\lambda) - F(\lambda)] = \frac{2\sqrt{T}}{\pi}\sum_{h=1}^{T-1}\frac{\sin \lambda h}{h}(r_h - \rho_h) - \frac{2\sqrt{T}}{\pi}\sum_{h=T}^{\infty}\frac{\sin \lambda h}{h}\rho_h \tag{11}$$

as a stochastic process over $[0, \pi]$. The monotonic transformation $u = G(\lambda)/G(\pi)$, where

$$G(\lambda) = 2\int_0^\lambda f^2(v)dv, \tag{12}$$

can be made from λ to u to obtain a process on $[0, 1]$. Then (11) is transformed to

$$Y_T(u) = \sqrt{T}\left[F_T\{G^{-1}[G(\pi)u]\} - F\{G^{-1}[G(\pi)u]\}\right], \tag{13}$$

and $Y_T(u)/[2\sqrt{\pi G(\pi)}]$ converges weakly to the Gaussian process with covariance function

$$k(u, v) = h(u, v) + q(u)q(v), \tag{14}$$

where

$$q(u) = u - F\{G^{-1}[G(\pi)u]\}, \tag{15}$$

and $h(u, v)$ is given by (2). If $B(u)$ is the Brownian bridge [that is, $\mathcal{E}B(u) = 0$ and $\mathcal{E}B(u)B(v) = h(u, v)$], then the limiting process above is $B(u) + q(u)X$, where X has the standard normal distribution $N(0, 1)$.

Test for a probability distribution: composite hypothesis. Suppose $\{F(x|\theta)\}$ is a family of probability cdf's indexed by the scalar parameter θ. If $\hat{\theta}$ is an estimator of θ, then

$$n\int_{-\infty}^{\infty}[F_n(x) - F(x|\hat{\theta})]^2 dF(x|\hat{\theta}) \tag{16}$$

can be used to test the null hypothesis that the cdf sampled belongs to the class $\{F(x|\theta)\}$. Let the density of $F(x|\theta)$ be $f(x|\theta)$; let

$$\mathcal{E}\left[\frac{\partial \log f(x|\theta)}{\partial \theta}\right]^2 = \mathcal{I}; \tag{17}$$

and let

$$m(u) = \frac{\partial \log F[F^{-1}(t|\theta)|\theta]}{\partial \theta}. \tag{18}$$

If $\hat{\theta}$ is an efficient estimator such as the maximum likelihood estimator, the process $\sqrt{n}\{F_n[F^{-1}(t|\theta)] - F[F^{-1}(t|\theta)|\hat{\theta}]\}$ converges weakly to a Gaussian process with covariance matrix $h(u,v) - m(u)m(v)/\mathcal{I}$.

In this paper we are interested in finding numerically the distributions of the limiting Cramér-von Mises statistic when the covariance function has the form $h(u,v) \pm r(u)r(v)$. In the case of a standardized spectral distribution the covariance function has this form with $r(u) = q(u)$ and the + sign; in the case of probability distributions with one estimated parameter the covariance function has this form with $r(u) = m(u)/\sqrt{\mathcal{I}}$ and the − sign.

2. Use of integral equations

First we sketch the general theory of finding the distribution of a quadratic integral of a continuous time parameter Gaussian stochastic process with a given covariance function by the example of the Cramér-von Mises criterion (1). Any function $k(u,v)$, $0 \le u,v \le 1$, that is symmetric in u and v, continuous, and square integrable (in one and both variables) can be expressed as

$$k(u,v) = \sum_{j=1}^{\infty} \frac{1}{\lambda_j} f_j(u) f_j(v), \tag{19}$$

where λ_j is an eigenvalue and $f_j(u)$ the corresponding normalized eigenfunction of the integral equation

$$f(u) = \lambda \int_0^1 k(u,v) f(v) dv \tag{20}$$

and

$$\int_0^1 f_i(u) f_j(u) du = \delta_{ij}, \tag{21}$$

where $\delta_{ii} = 1$ and $\delta_{ij} = 0$, $i \neq j$. If $k(0,0) = k(1,1) = 0$, then $f_i(0) = f_i(1) = 0$. If $k(u,v)$ is the covariance function of a stochastic process $Z(u)$, $0 \leq u \leq 1$, then $k(u,v)$ is positive definite and $\lambda_j > 0$. The series (19) converges absolutely and uniformly in the unit square. If $Z(u)$ is Gaussian and $\mathcal{E}Z(u) = 0$, then it has the representation

$$Z(u) = \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j}} X_j f_j(u), \tag{22}$$

where $X_1, X_2, \cdots$ are independent $N(0,1)$ variables. With probability 1,

$$\int_0^1 Z^2(u)du = \int_0^1 \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_i}} X_i f_i(u) \frac{1}{\sqrt{\lambda_j}} X_j f_j(u) du = \sum_{j=1}^{\infty} \frac{1}{\lambda_j} X_j^2. \tag{23}$$

The characteristic function of (23) is

$$\mathcal{E} \exp\left(it \sum_{j=1}^{\infty} \frac{X_j^2}{\lambda_j}\right) = \prod_{j=1}^{\infty} \left(1 - \frac{2it}{\lambda_j}\right)^{-1/2}. \tag{24}$$

The function $D(\lambda) = \prod_{j=1}^{\infty}(1 - \lambda/\lambda_j)$ is known as the Fredholm determinant of the integral equation (20).

The Brownian bridge with covariance function $h(u,v) = \min(u,v) - uv$ has the representation (22) with $\lambda_j = \pi^2 j^2$ and $f_j(u) = \sqrt{2}\sin j\pi u$. The characteristic function (24) is $(\sin\sqrt{2it}/\sqrt{2it})^{-\frac{1}{2}}$ [Anderson and Darling (1952)].

Now consider $k(u,v) = h(u,v) + q(u)q(v)$. We define the Fourier coefficients

$$\begin{aligned} \alpha_j &= \int_0^1 q(u) f_j(u) du \\ &= \frac{2\sqrt{2}}{G(\pi)} \int_0^{\pi} \sin\left[j\pi \frac{G(\lambda)}{G(\pi)}\right] \left[\frac{G(\lambda)}{G(\pi)} - F(\lambda)\right] f^2(\lambda) d\lambda. \end{aligned} \tag{25}$$

Then $q(u) = \sum_{j=1}^{\infty} \alpha_j f_j(u)$. The process $B(u) + Xq(u)$ has the representation

$$B(u) + Xq(u) = \sum_{j=1}^{\infty} \left(\frac{X_j}{\sqrt{\lambda_j}} + \alpha_j X\right) f_j(u), \tag{26}$$

and the Cramér-von Mises criterion has the representation

$$S = \int_0^1 [B(u) + Xq(u)]^2 du \tag{27}$$

$$= \int_0^1 \left[\sum_{j=1}^{\infty}\left(\frac{X_j}{\sqrt{\lambda_j}} + \alpha X\right) f_j(u)\right]^2 du$$

$$= \sum_{j=1}^{\infty}\left(\frac{X_j}{\sqrt{\lambda_j}} + \alpha X\right)^2.$$

Alternatively, there is a representation

$$k(u,v) = \sum_{j=1}^{\infty} \frac{1}{\nu_j} g_j(u) g_j(v), \tag{28}$$

where $k(u,v)$ is given by (14) and ν_j and $g_j(u)$ are the eigenvalues and eigenfunctions associated with the integral equation

$$g(u) = \nu \int_0^1 k(u,v) g(v) dv. \tag{29}$$

The Fredholm determinant of (29) is

$$D^*(\nu) = \frac{\sin\sqrt{\nu}}{\sqrt{\nu}}\{1 - \nu^2 \int_0^1 \int_0^1 c(u,t;\nu) q(u) q(t) du dt - \nu \int_0^1 q^2(u) du\}, \tag{30}$$

where

$$c(u,t;\nu) = \begin{cases} -\frac{1}{\sqrt{\nu}\sin\sqrt{\nu}} \sin(\sqrt{\nu}u)\sin[\sqrt{\nu}(t-1)], & u \le t, \\ -\frac{1}{\sqrt{\nu}\sin\sqrt{\nu}} \sin[\sqrt{\nu}(u-1)]\sin(\sqrt{\nu}t), & u \ge t. \end{cases} \tag{31}$$

The eigenvalues of $k(u,v)$ are the values of ν ($\neq \pi^2 j^2$) that satisfy $D^*(\nu) = 0$. The function $c(u,t;\nu)$ is the *resolvent or resolving kernel* of the kernel $\min(u,t) - ut$. The characteristic function of S is $[D^*(it)]^{-\frac{1}{2}}$.

Let $X_j/\sqrt{\lambda_j} + \alpha_j X = Y_j, \quad j = 1, 2, \cdots$. Then (27) is

$$S = \sum_{j=1}^{\infty} Y_j^2. \tag{32}$$

The Y_j are normally distributed with $\mathcal{E}Y_j = 0$, $\mathcal{E}Y_j^2 = 1/\lambda_j + \alpha_j^2$, and $\mathcal{E}Y_iY_j = \alpha_i\alpha_j$, $i \neq j$. Then S can be approximated by a finite sum $S_N = \sum_{j=1}^{N} Y_j^2$. The difference $S - S_N$ has expectation

$$\mathcal{E}(S - S_N) = \mathcal{E}\sum_{j=N+1}^{\infty} Y_j^2 = \sum_{j=N+1}^{\infty}\left(\frac{1}{\lambda_j} + \alpha_j^2\right), \tag{33}$$

which can be made arbitrarily small by taking N sufficiently large. Hence, as $N \to \infty$, the distribution of S_N converges to the distribution of S and the characteristic function of S_N approaches the characteristic function of S.

Let Y_N be the N-vector with Y_j as the jth component. The covariance matrix of Y_N is $\mathcal{E}Y_N Y_N' = \Lambda_N + \alpha_N \alpha_N'$, where Λ_N is the diagonal matrix with $1/\lambda_j$ as the jth diagonal element and α_N is the vector with α_j as the jth component. Then the characteristic function of S_N is

$$\mathcal{E}e^{itY_N' Y_N} = |I_N - 2it(\Lambda_N + \alpha_N \alpha_N')|^{-1/2} = \prod_{j=1}^{N} (1 - 2it\phi_{jN})^{-1/2}, \tag{34}$$

where ϕ_{jN} ($\phi_{1N} \geq \phi_{2N} \geq \cdots \geq \phi_{NN}$) is the jth characteristic root of $\Lambda_N + \alpha_N \alpha_N'$, that is, the jth zero of

$$|\Lambda_N + \alpha_N \alpha_N' - \phi I_N| = |\Lambda_N - \phi I_N| \psi(\phi), \tag{35}$$

where

$$\psi(\phi) = 1 + \sum_{j=1}^{N} \frac{\alpha_j^2}{\frac{1}{\lambda_j} - \phi} \tag{36}$$

for $\phi \neq 1/\lambda_j$. If $\alpha_j \neq 0$ (as is the case for the MA(1) and AR(1) tests tabulated below), $1/\lambda_j$ is not a zero of (35). Let

$$\begin{aligned} D_N^*(\nu) &= |I_N - \nu(\Lambda_N + \alpha_N \alpha_N')| \\ &= \prod_{i=1}^{N} \left(1 - \frac{\nu}{\lambda_i}\right) \left\{1 - \nu^2 \sum_{j=1}^{N} \frac{\alpha_j^2}{\lambda_j - \nu} - \nu \sum_{j=1}^{N} \alpha_j^2\right\}. \end{aligned} \tag{37}$$

Since $\sum_{j=1}^{\infty} \lambda_j^{-1} < \infty$ and $\sum_{j=1}^{\infty} \alpha_j^2 = \int_0^1 q^2(u)du < \infty$, $D_N^*(\nu)$ converges to $D^*(\nu)$, the Fredholm determinant, as $N \to \infty$. The characteristic function of S is $1/\sqrt{D^*(2it)}$.

When the explicit form (30) of the Fredholm determinant is intractable or cannot be inverted, it can be approximated by $D_N^*(\nu)$ given by (37). For this we turn to the numerical evaluation of the ϕ_{jN}, the characteristic roots of $\Lambda_N + \alpha_N \alpha_N'$ and the zeros of (35). These approximate the reciprocals of the first N eigenvalues of $k(u, v)$. The first derivative of $\psi(\phi)$ is

$$\psi'(\phi) = \sum_{j=1}^{N} \frac{\alpha_j^2}{(\frac{1}{\lambda_j} - \phi)^2} > 0. \tag{38}$$

As $\phi \to 1/\lambda_j$ from above, $\psi(\phi) \to -\infty$, and as $\phi \to 1/\lambda_j$ from below, $\psi(\phi) \to \infty$. Since $\psi(\phi)$ is continuous except at $\phi = 1/\lambda_j$, $j = 1, \cdots, N$, there is one root in the interval $(1/\lambda_j, 1/\lambda_{j-1})$, $j = 1, \cdots, N$, where

$$1/\lambda_0 = \sum_{j=1}^{N} \alpha_j^2. \tag{39}$$

3. Calculation of the distributions of test statistics

In this section the theory given above is employed to provide tests for two time series models, the MA(1) and AR(1) processes. Asymptotic points are given with the tests; the procedure for calculating these points follows closely the methods employed by Stephens (1976) for obtaining asymptotic points for goodness-of-fit statistics in connection with tests for distributions. We first find approximations to the ν_j, or more precisely, the values $\omega_j = 1/\nu_j$, in the representation (28). These are the zeros ϕ of (36). They were found by two methods, one used to check the other, and in each case asymptotic points were calculated.

Method of Fourier coefficients. Recall that when the covariance of the Gaussian process is $h(u, v)$ in (2), the eigenvalues are $\lambda_j = \pi^2 j^2$ and the eigenfunctions are $f_j(u) = \sqrt{2}\sin(\pi j u)$. In order to apply the results of Section 2 above, we need values α_j as defined by (25), where the functions $q(u)$ depend on the asymptotic Gaussian process for the time series model. Anderson (1993) gave explicitly the functions $F(\lambda)$ and $G(\lambda)$ for the MA(1) and AR(1) models. The values of α_j defined by the second expression in (25) were calculated by numerical integration for $j = 1, \cdots, 40$; they are all nonzero, but rapidly become very small (of the order of 10^{-4} for $j = 40$). Note that if the first expression in (25) is approximated by $\sqrt{2}\sum_{i=0}^{n} q(i/n)\sin(j\pi i/n)/n$, obtained by straightforward application of the trapezoidal rule, the Fast Fourier Transform can be used to obtain values of α_j for many j. For this calculation, $q(u)$ given by (15) must be found. This was done by calculating a table of values of $G(\lambda)$ at intervals small enough to obtain a first approximation to $G^{-1}[G(\pi)u]$ by linear interpolation;

the approximation was then refined by Newton's method.

The zeros of (36) can be found by search since (38) indicates that $\psi(\phi)$ is always increasing between vertical asymptotes at $1/\lambda_j$. The ω_j for $j = 41, \ldots, 100$ were approximated by observing that $\omega_j/j^2 \to k$, a constant which is determined from the last calculated values.

Method of discretization With modern computing resources we can find approximate values of ω_j by creating a matrix K with entries $k_{ij} = k(u_i, v_j)$; $k(u, v)$ is given by (14) and $u_i = i/(n+1)$ and $v_j = j/(n+1)$, for $i, j = 1, \ldots, n$. The characteristic roots and vectors of K then approximate ω_j and the corresponding eigenfunctions $g_j(u)$ of (29). This computation was done for $n = 50, 100$ and 200.

Calculation of asymptotic points When the set of n values of ω_j is found, the distribution of S given by (23) or (32) is approximated by the distribution of

$$T = \sum_{j=1}^{n} \omega_j X_j^2 + c\,, \tag{40}$$

where X_j are independent standard normals, and c is a constant, obtained from the fact that

$$\mathcal{E}S = \int_0^1 k(u,u)du = \mathcal{E}T = \sum_{j=1}^{n} \omega_j + c. \tag{41}$$

The integral $\int_0^1 k(u,u)du = 1/6 + \int_0^1 q^2(u)du$, and the last integral is found numerically, thus determining c. It might be thought that c should be replaced by a random variable, say ω_∞, but the difference in $\mathcal{E}T$ and $\sum_{j=1}^{n} \omega_j$ is sufficiently small that the variance of ω_∞ is negligible.

Given $\omega_1, \cdots, \omega_n$, T is a sum of c and a weighted sum of χ^2's, each with 1 degree of freedom. Then $\Pr\{T - c \leq x\}$ can be found by Imhof's method (for which computer programs are available). For selected values of the probabilities the values of x were computed for Tables 1 and 2. It was found that when x for a given probability was plotted against $1/n$ $(n = 50, 100, 200)$ in the second method the plot was linear and hence could be extrapolated to $1/n = 0$ to give the final percentage points at this probability. The values obtained in this way agreed to those by the first method to

4 decimal places in the upper tail (which is the tail used for significance points) and differed by only 1 in the fourth decimal place in the lower tail. If only one method is to be used, the method of Fourier coefficients seems to be preferable.

4. Test for the moving average model of order 1

For this model, the time series y_t is given by $y_t = u_t + \alpha u_{t-1}$, where the u_t are uncorrelated with mean zero and variance σ^2. Let the autocovariance be $\mathcal{E}y_t y_{t+h} = \sigma(h)$; the autocorrelation is $\rho_h = \sigma(h)/\sigma(0)$. Then $\rho_0 = 1$, $\rho_1 = \rho_{-1} = \rho = \alpha/(1+\alpha^2)$, and $\rho_h = 0$ for $h \neq -1$, 0 or 1. The test below is for a moving average model with α (or equivalently ρ) known. Suppose $y_1, y_2, \ldots, y_T$ is a sample of T successive values of y_t. The sample autocovariances are given by (7), and the sample autocorrelations are $r_h = r_{-h} = c_h/c_0$, $h = 0, 1, \ldots, T-1$.

The test statistic The Cramér-von Mises criterion for a time series process yields a test statistic with computing formula [Anderson, (1993)]

$$Q = \frac{T}{8\pi^4 G^2(\pi)} \sum_{r=-T}^{T} \left\{ \sum_{g=1}^{T-1} \frac{(r_g - \rho_g)(\rho_{r+g} - \rho_{r-g})}{g} \right\}^2, \tag{42}$$

where $G(\cdot)$ depends on the process. In calculating Q from (11) the second sum on the right-hand side of (11) has been omitted because it converges to zero. For the MA(1) process, $G(\pi) = (1+2\rho^2)/(2\pi)$ [Anderson, (1993)]. Also since ρ_h takes values only for $h = -1, 0$ and 1, the formula (41) simplifies considerably for this process: for a given r, the sum over g can be divided into two sums for g from $-r-1$ to $-r+1$ and for g from $r-1$ to $r+1$. The test for the MA(1) process with known α, that is, with known ρ, then consists of the following steps:

1. Calculate the r_g for $g = 1, \ldots, T-1$;
2. Calculate Q;
3. Refer Q to the asymptotic points in Table 1 for the corresponding ρ. Reject the MA(1) model at level p if Q exceeds the percentage point given for level p.

The points are asymptotic, but extensive Monte Carlo studies indicate that they can be used with good accuracy for $T > 50$, a reasonable length for most time series.

5. Test for the autoregressive model of order 1

For this time series, $y_t = \rho y_{t-1} + u_t$, where the u_t are as defined above. Then, $\rho_h = \rho_{-h} = \rho^h$, and $G(\pi) = (1 + \rho^2)/\{2\pi(1 - \rho^2)\}$. Given a sample of length T, the test for AR(1) with known ρ follows the three steps given in the previous section, but Q is now referred to the asymptotic points in Table 2 for the corresponding ρ. Reject the AR(1) model at level p if Q exceeds the percentage point given for level p.

Here again, the asymptotic points will be adequate for $T > 50$. For this process, Q is more complicated to calculate and it is important to keep to the correct limits for both r and g. For both models, a FORTRAN program to calculate the test statistic and also to look up the appropriate table, is available from the second author.

The authors are very grateful to Dr. Richard Lockhart for valuable assistance with the computational procedures in this paper.

References

Anderson, T. W. (1993). Goodness of fit tests for spectral distributions, *The Annals of Statistics* **21**, 830-847.

Anderson, T. W., and D. A. Darling (1952). Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes. *Ann. Math. Statist.* **23**, 193-212.

Stephens, M. A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters, *Annals of Statistics,* **4**, 357-369.

Table 1.: Asymptotic Points For MA(1) Test

$\rho \setminus p$	0.001	0.005	0.01	0.025	0.05	0.1	0.15	0.25
0.0	0.0170	0.0218	0.0248	0.0304	0.0366	0.0460	0.0543	0.0703
0.1	0.0170	0.0218	0.0248	0.0304	0.0367	0.0462	0.0545	0.0706
0.2	0.0171	0.0219	0.0250	0.0306	0.0369	0.0466	0.0551	0.0715
0.3	0.0172	0.0221	0.0252	0.0309	0.0373	0.0471	0.0558	0.0726
0.4	0.0173	0.0222	0.0253	0.0310	0.0375	0.0474	0.0561	0.0731
0.5	0.0172	0.0221	0.0252	0.0310	0.0374	0.0472	0.0559	0.0727

Table 1.: Asymptotic Points For MA(1) Test cont.

$\rho \setminus p$	0.5	0.75	0.85	0.9	0.95	0.975	0.99	0.995	0.999
0.0	0.1189	0.2094	0.2841	0.3473	0.4614	0.5806	0.7435	0.8694	1.1679
0.1	0.1198	0.2118	0.2878	0.3522	0.4684	0.5898	0.7557	0.8839	1.1879
0.2	0.1221	0.2172	0.2961	0.3630	0.4837	0.6099	0.7821	0.9152	1.2308
0.3	0.1244	0.2221	0.3034	0.3722	0.4966	0.6264	0.8037	0.9408	1.2657
0.4	0.1253	0.2238	0.3057	0.3751	0.5004	0.6313	0.8100	0.9481	1.2755
0.5	0.1244	0.2219	0.3028	0.3714	0.4951	0.6245	0.8011	0.9376	1.2611

Table 2.: Asymptotic Points For AR(1) Test

$\rho \setminus p$	0.001	0.005	0.01	0.025	0.05	0.1	0.15	0.25
0.0	0.0170	0.0218	0.0248	0.0304	0.0366	0.0460	0.0543	0.0703
0.1	0.0170	0.0218	0.0248	0.0304	0.0367	0.0462	0.0545	0.0706
0.2	0.0171	0.0219	0.0250	0.0306	0.0370	0.0467	0.0552	0.0717
0.3	0.0172	0.0221	0.0252	0.0310	0.0375	0.0474	0.0562	0.0734
0.4	0.0174	0.0224	0.0256	0.0315	0.0382	0.0484	0.0576	0.0755
0.5	0.0176	0.0227	0.0259	0.0320	0.0389	0.0495	0.0590	0.0776
0.6	0.0178	0.0230	0.0263	0.0326	0.0396	0.0506	0.0603	0.0796
0.7	0.0180	0.0233	0.0267	0.0330	0.0402	0.0514	0.0614	0.0812
0.8	0.0181	0.0235	0.0269	0.0334	0.0407	0.0520	0.0622	0.0822
0.9	0.0182	0.0236	0.0271	0.0336	0.0409	0.0524	0.0626	0.0828

Table 2.: Asymptotic Points For AR(1) Test *cont.*

$\rho \setminus p$	0.5	0.75	0.85	0.9	0.95	0.975	0.99	0.995	0.999
0.0	0.1189	0.2094	0.2841	0.3473	0.4614	0.5806	0.7424	0.8694	1.1679
0.1	0.1198	0.2119	0.2880	0.3525	0.4688	0.5904	0.7567	0.8848	1.1891
0.2	0.1226	0.2188	0.2988	0.3666	0.4890	0.6168	0.7923	0.9263	1.2462
0.3	0.1267	0.2288	0.3140	0.3864	0.5169	0.6533	0.8393	0.9831	1.3242
0.4	0.1316	0.2399	0.3307	0.4079	0.5470	0.6924	0.8905	1.0439	1.4072
0.5	0.1364	0.2505	0.3465	0.4280	0.5749	0.7284	0.9378	1.0997	1.4831
0.6	0.1406	0.2595	0.3596	0.4447	0.5980	0.7581	0.9765	1.1453	1.5469
0.7	0.1438	0.2662	0.3694	0.4570	0.6150	0.7800	1.0051	1.1790	1.5901
0.8	0.1460	0.2707	0.3758	0.4651	0.6262	0.7943	1.0238	1.2010	1.6213
0.9	0.1472	0.2731	0.3792	0.4695	0.6321	0.8020	1.0337	1.2127	1.6388

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 488	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Modified Cramer-von Mises Goodness-of-Fit Criterion for Time Series		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) T.W. Anderson and M.A. Stephens		8. CONTRACT OR GRANT NUMBER(s) N0025-92-J-1264
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305-4065		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 111		12. REPORT DATE 17 January 1994
		13. NUMBER OF PAGES 14
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/ DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Goodness of Fit, Time Series, Mahalanobis distance, stationary stochastic process, spectral distributions.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See Reverse Side

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

In this paper we consider somewhat analogous quadratic forms in normal variables when the dimensionality is infinite. Then the quadratic forms are distributed as infinite weighted sums of χ^2-variables. These come about as goodness-of-fit criteria for a hypothesis that a cumulative distribution function is a specified one or that two cdf's are the same. Such criteria also arise for goodness-of-fit tests for standardized spectral distributions.

As examples, we give tables of the distribution of the criterion for testing the hypothesis that a stationary stochastic process is a given moving average process of order 1 and for testing the hypothesis that it is a specified autoregressive process of order 1. Two methods are described for calculating the distribution. Either method is appropriate for calculating the distribution of the criterion for testing the hypothesis that a process is a stationary process whose standardized spectral density or distribution is a specified one.